

## **WP3**

### **Information Platform**

Lead contractor: Faculty of Sciences University of Lisbon

**Objectives.** This task develops the Epidemic Marketplace, Epiwork's information integration platform, the environment where epidemiological data can be stored, managed and made available to investigators, fostering collaboration. The objectives of Epiwork where the Epidemic Marketplace will have a direct impact are: (1) the development of large scale, data driven computational models endowed with a high level of realism and aimed at epidemic scenario forecast; (2) the design and implementation of original data-collection schemes motivated by identified modelling needs, such as the collection of real-time disease incidence, through innovative Internet and ICT applications; (3) the set up of a computational platform for epidemic research and data sharing that will generate important synergies between research communities and states..

### **Progress towards objectives**

The work package activities have proceeded and the deliverables planned for the first year have been completed on time, despite the difficulties in hiring the people initially envisioned to conduct the development tasks at the University of Lisbon, Faculty of Sciences. We expect to recover the delays that this caused to the project and use the non-spent personnel budget in the coming years. The two deliverables have been completed on time, despite these difficulties, albeit the functionalities that we have at this time are not as comprehensive as expected. However, we decided that it would be better to make the existing prototype available to the consortium than delaying its announcement.

Task 1: Data collection. Realistic simulations of epidemic processes crucially depend on the availability of datasets describing human behaviour and pathogen-host interactions. Datasets include population movement data, social and behavioural data, health related data, geographic data, detailed geo-temporal epidemic incidence and immunization data, pathogen evolution and multi-strains circulation data. Data can come from a variety of different sources, including hospital records, country statistics, Web content, and others. It can range from a global scale, such as the worldwide air transportation infrastructure, down to the detailed description of individual activities at a minute-by-minute scale. This task is creating a catalogue of databases of epidemiological data across Europe, with extensive meta-data describing the main characteristics of the available information sources. The data collection activity by the consortium started at M13, after the deployment of first functional prototype and its released to the consortium (**Deliverable D3.2**). In the first semester, we started with some initial experiments, including:

- Initial catalogue design, i.e., initiated the conceptual model for managing large datasets that is now under development.
- Design and implementation of a data collection prototype based on the twitter API and development platform (flu-related tweets).

In the second semester, we initiated the development of a second prototype of a data collector and the development of an initial prototype of the Epidemic Marketplace services.

Task 2 : Meta-model design. The Epidemic Marketplace Platform supports the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. Each dataset will come with a metadata file, signalling the date of submission, the last update, the source of the dataset, a basic profile (e.g., transportation network – Origin-Destination matrix), and a more thorough description of the dataset and the classification used. Work in this task, most of it detailed in **Deliverable D3.1**, included:

- A review of meta-modelling techniques and existing standards.
- Characterisation of the data sources most commonly used in epidemiological studies.
- Initial design of the epidemic meta-data catalogue.

In October 2009, we initiated the development of an Epiwork meta-data editor for epidemic datasets, following the policies outlined in deliverable D3.1.

The activity on Task 3 : Epidemic Marketplace Platform has progressed towards the implementation a platform based on the integration of grid technology and publicly available services and software on the web to support the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. Researchers can use and contribute to the Marketplace in several different ways. They can: (1) use it as a catalogue of data sources containing the metadata describing existing databases; (2) view, download, tag, and comment on the available resources; (3) provide compliant datasets and relevant information; (4) use it as a forum where to publish information about their own data, seek modellers to collaborate with, share and distribute their new findings. Progress in this task in the first year included:

- Definition of the general architecture of the Epidemic Marketplace. This has involved joint discussions with WP5 on how data collected data by IMS in different countries is aggregated in datasets and uploaded to the Epidemic Marketplace. Privacy and anonymization issues have emerged.
- Infrastructure design and identification of equipment to be acquired for Epiwork.
- Installation of the hardware and base software (OS)
- Configuration and implementation of fault tolerance support for the epidemic marketplace: backup and data replication policies
- EM services installation, including Epidemic Marketplace's Repository main components: Fedora Commons ([www.fedora-commons.org](http://www.fedora-commons.org)) and Muradora (<http://www.muradora.org/muradora>). The Repository is now accessible to authenticated users from <http://epiwork.di.fc.ul.pt/muradora/>
- Deployment of a first prototype version of the Epidemic Marketplace Forum, based on phpBB, <http://www.phpbb.com/>.

The first operational prototype of the Epidemic Marketplace was presented in the November 2009 meeting in Torino, and made available to Consortium members since then from <http://epiwork.di.fc.ul.pt>. In January 2010, we initiated the deployment of a new version of the EM base software, to be based in Fedora Commons version 3.0 and

the Drupal Content Management System. This new version will substitute the current, based of Fedora Commons v2.0 and Muradora.

- Discussions with partners involved in WP3 on how to identify relevant datasets to the catalogue and strategies and incentives for populating the Epidemic Marketplace.

The activity on Task 4: Evaluation and monitoring of the use of the catalogue and collaboration services was not scheduled to start in the first year. Evaluation work is to take place only once the information platform is deployed.

## **Publications and working papers**

- Luis F. Lopes, João M. Zamite, Bruno C. Tavares, Francisco M. Couto, Fabrício Silva and Mário J. Silva. *Automated Social Network Epidemic Data Collector*. Inforum, September 2009.
- Luís Filipe Lopes, Fabrício Silva, Francisco Couto, Mário J. Silva, *Meta-model initial specification, catalogue of relevant data, platform requirements*. Deliverable D3.1 of the EPIWORK Project. September, 2009.
- Mário J. Silva, Fabrício A. B. da Silva, Luís Filipe Lopes, Francisco M. Couto. *Building a Digital Library for Epidemic Modelling*. Invited Paper. ICDL 2010. The 3rd International Conference on Digital Libraries. February, 2010.
- Mário J. Silva, Fabrício A. B. da Silva, Luís Filipe Lopes, Francisco M. Couto. *Prototype of the Epidemic Marketplace Platform with an initial set of epidemiological databases integrated available to project participants*. Deliverable D3.2 of the EPIWORK Project. January, 2010.